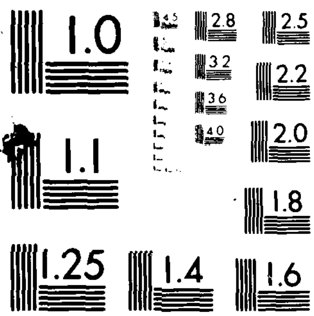
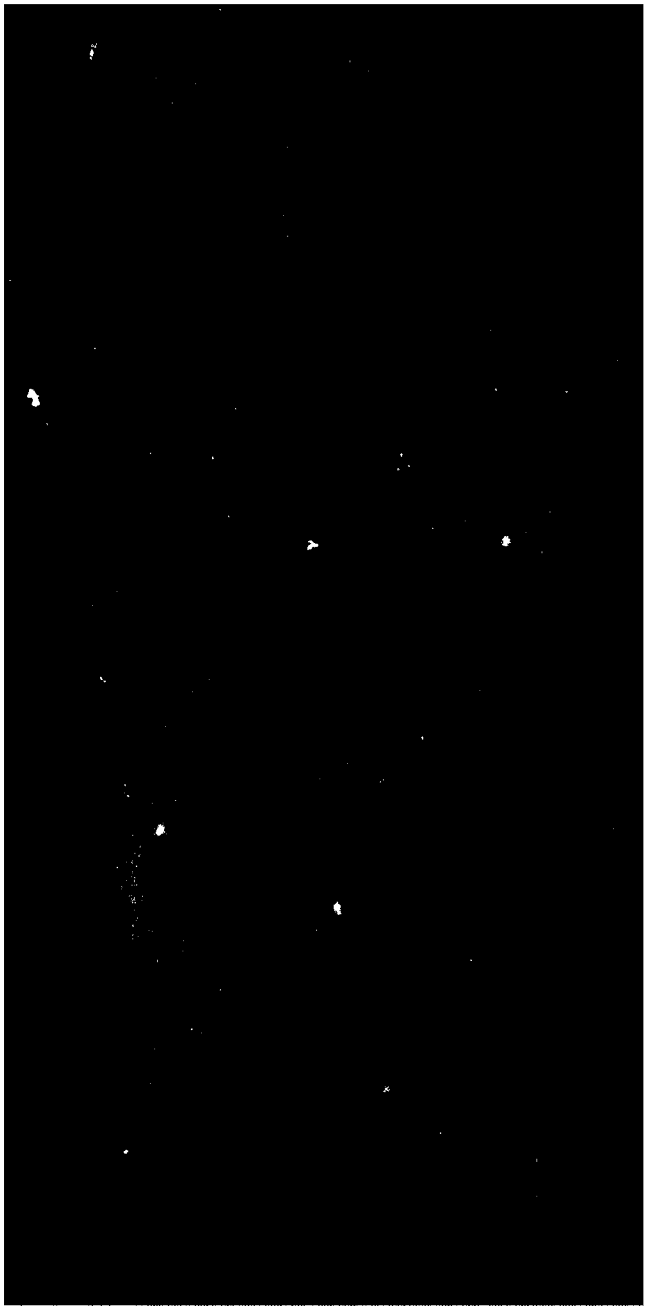
AD-A112 949 CLEMSON UNIV SC DEPT OF MATHEMATICAL SCIENCES F/G 12/1
THE BASKET METHOD FOR SELECTING BALANCED SAMPLES. PART II. APPL--ETC(U)
DEC 81 K T WALLENIUS N00014-75-C-0451
UNCLASSIFIED N132 NPS-NCAR-81-1 NL

END
DATE
FILMED
5 82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A



(12)

THE BASKET METHOD FOR SELECTING BALANCED SAMPLES - PART II: APPLICATIONS TO PRICE ESTIMATION

BY

K. T. WALLENIUS*

DEPARTMENT OF MATHEMATICAL SCIENCES

CLEMSON UNIVERSITY

TECHNICAL REPORT #377
DECEMBER 1981
N 132

*The methodology upon which this descriptive report is based was developed under Contract N00014-75-C-0451 with the Office of Naval Research. Support for the preparation of this document was supplied by ONR through the Navy Center for Acquisition Research (NCAR), Naval Post Graduate School, Monterey, California where the author is on sabbatical leave during academic year 1981-82. This report also appears as NCAR Report 81-1.

The Basket Method for Selecting Balanced Samples

Part II: Applications to Price Estimation

ABSTRACT

The "Basket Method" of sampling, a tool designed to achieve statistically balanced samples, is described in intuitive terms. Special reference is made to applications in price analysis where experience has demonstrated the practicality of the technique. The intent is to provide an overview of what the system is intended to do and how it does it in order to assist price analysts and negotiators expedite proposal processing while maintaining acceptable levels of risk. Guidelines and examples are given for implementing a statistical pricing program tailored to local conditions. Underlying theory and documented computer codes are provided separately in Part I and Part III, respectively.

TABLE OF CONTENTS

Introduction and Background

Current defense acquisition practices often involve situations in which Department of Defense (DOD) agencies must deal with a sole source in buying commodities or services. A few examples will demonstrate why it's sometimes necessary, even desirable, for the buyer to negotiate with a single supplier. In each situation, the basket method has proven its practical value as a technique for expediting required price analyses through the use of statistical inference.

(a) Change orders: After a prime contract is awarded and production begins, design changes are often necessitated by a change in performance requirements requested by the government or by unforseen technical problems which inevitably seem to crop up. Each design change requires a modification to the prime contract called a change order. The two parties must negotiate a price for each change - i.e. no competition.

(b) Provisioned items (spare parts, special test equipment, operations and repair manuals, etc.): The cost of these items is generally not an element of prime contract competition. Just how many spare parts will be needed is a function of product quality as well as reliability/maintainability policies. Uncertainties which exist at the time the prime contract is awarded will be greatly reduced as production proceeds, thus providing a better basis for accurate analysis when the time comes to negotiate prices for these items. Since the prime contractor has already performed costly engineering, tooling and manufacturing functions to produce the total system, he would certainly win any competition for the production of spare parts. Thus, the cost of such items must be negotiated from a sole source without the benefit of competition.

(c) Major modification or repair of a system subsequent to production: Again, because of his prior investment and experience, the producer of the system has such an advantage over competitors that there is no effective competition.

In each of these examples, the sole source prepares a proposal for every requested change, provisioned item, or modification. Defense Acquisition Regulations (DAR) require analysis of the cost proposed and negotiation of these costs for each proposal by some cognizant agency, usually a group of government employees assigned to perform such functions at the contractor's facility. The volume of work thus generated and the amount of money involved is quite substantial. This volume of work coupled with insufficient numbers of government analysts and negotiators leads to large backlogs of unprocessed proposals. At one particular aircraft plant visited, there were some 625 unprocessed proposals totaling nearly half a billion dollars. At a shipyard the figures were 2,500 at .2 billion. To perform a really thorough analysis and patient negotiation would have taken the small on-site staff several months of round-the-clock work. And, during this time, the contractor would be submitting proposals for other jobs so the backlog would still exist. Thus, there is tremendous pressure on analysts to expedite their work even though it is generally recognized that hurried analysis and negotiation can result in costly overpayment since quickness generally works against thoroughness and accuracy. Generally, backlogs develop because quickness is sacrificed for thorough analysis.

On the other side of the coin, unprocessed proposals can result in extra expense for the contractor. In many situations involving ongoing production or repair processes, the proposed work goes on before the proposal is analyzed and negotiated to avoid expensive delay and disruption costs. Despite partial advances (progress payments), the contractor may have to borrow working capital to cover funds tied up in the backlog thus suffering capital costs.

It is generally recognized that it is in the best interest of both parties to expedite the processing of the proposals in accordance with DAR provisions. If government price analysts could somehow look into a crystal ball and foretell what the total negotiated price would be if each proposal were carefully analyzed and negotiated, and if such a practice were legal, the conflicting goals of thorough analysis and quickness could be reconciled. Statistical inference can be thought of as a middle ground between business as usual and crystal ballery: By selecting a suitable sample of proposals from the backlog, carefully analyzing the accuracy of proposed costs and negotiating a price for each sampled proposal, the resulting data can be extrapolated to estimate what the results would have been had every proposal received the same treatment.

In this paper we discuss such a sampling procedure called the "Basket Method" which was specifically developed to help expedite the negotiation process. It is based on sound statistical principles and has been successfully used by various organizations in aircraft procurement, ship procurement and repair, and missile procurement. This paper, including its appendix, is intended to be a self-contained introduction to the "Basket Method." The body of the text describes how the method works and the appendix contains detailed instructions for using computer programs available to COPPER IMPACT subscribers. A general overview, including a brief statistics tutorial written for undergraduate mathematics students can be found in reference (2). A more rigorous mathematical treatment is given in reference (3). Documented computer codes are contained in reference (4) so that users with unusual requirements can modify the source program to suit their specific needs.

This paper is organized in six parts:

A. Sampling and Estimation: Fundamental concepts.

B. The "Basket Method": Basic ideas.

C. Perspective Analysis: Tailoring the sampling procedure to local conditions.

D. *Fine Tuning:* Optional use of multilevel attributes.

E. Additional Guidelines.

F. Appendix: Use of computer programs.

Part A

Sampling and Estimation: Fundamental Concepts

These are 4 basic components common to all sampling and estimation problems like sampling from a proposal backlog:

1. A well defined POPULATION of units. We will use the letter N to denote the number of units in the population. Associated with each unit in the population is an unknown characteristic of interest which can be learned if the unit is selected for examination. We will use the letter Y to denote the numerical value of this characteristic. The total of all the Y values in the population, which we will denote by T, is the quantity we wish to estimate. Besides this unknown quantity Y, there may be auxiliary information about each unit which is accessible prior to selecting the sample and estimating T.

2. A SAMPLING PLAN. This is a description of how the sample will be selected from the population. We will use the letter n to denote the number of units in the sample.

3. An ESTIMATION RULE. This is a formula which uses information learned from the sample to estimate T.

4. A statement about the PROBABLE ERROR associated with the estimated total. By "error" we mean the difference between the true value of T for the population of N units and its estimated value based on the sample of size n. Error-free estimation cannot be expected unless n=N, that is, the entire population is "screened". Generally speaking, one controls the probable error by selecting a suitably large sample size n. Selecting n is usually a trade-off between the cost of processing the sample and a tolerable probable error.

In the present scenario, the units of the POPULATION are price proposals. The population will generally be only a portion of the backlog since it may be inappropriate to subject all the proposals in the backlog to statistical estimation. For example, certain provisions in the DAR have been interpreted as limiting proposals in the sample to under $100,000. Other considerations in forming the population are discussed in Part E. The unknown characteristic of interest Y is the negotiated price. Auxiliary information includes the proposed price, which we denote by the letter X, as well as other features such as contractual type (fixed price, cost plus fix fee, etc.), type of job (repair, change order, spares, etc.) and description of costs (labor hours, material costs, overhead, etc.). Up to this point, the scenario is typical of most sampling situations but care must be exercised in selecting a SAMPLING PLAN and ESTIMATION RULE because of an atypical ingredient in the scenario: the potential for gamesmanship.

Statistical inference is often described as a game between the statistician and "nature". We can imagine "nature" as a player who sets the values of the characteristics which are unknown to the statistician. The statistician obtains partial information about the characteristics by sampling and uses that information to make inferences about the unsampled units. Inference errors can be viewed as "losses" to the statistician, but we do not generally view "nature" as a hostile entity who benefits through the statistician's estimation misfortunes. In our problem, however, if we assume the contractor has a fairly good idea of what the negotiated outcome of each proposal should be, and if we define the "state of nature" for each proposal as the dollar decrement X-Y, then the contractor, by setting the proposed price X, <u>controls</u> the state of nature. While we are not saying it's in the contractor's best interest to try to selectively pad proposals in order to exploit a systematically vulnerable sampling and estimation procedure, it would be a mistake to use such a vulnerable procedure thereby inviting gamemanship attempts. Since both parties to the negotiation must enter into a binding agreement to abide by the results of applying statistical methods, all aspects of the sampling and estimation process must be disclosed in advance. With this necessary advanced knowledge, a clever contractor could exploit vulnerabilities of many standard statistical techniques as shown in reference (2). The "Basket Method" takes the competitive edge away from either side and assures a statistically unbiased determination of total cost.

While sample selection and estimation can be accomplished automatically by use of a computer, in no way does the computer diminish the role of the price analyst. Indeed, the importance of sound analysis and negotiation is magnified since, as will be seen, decrement factors negotiated in the sample are applied to the unsampled portion of the backlog. In addition to analyzing and negotiating the sampled proposals, the price analyst's experience with local business practice is crucial in setting up and monitoring a statistical sampling program and in providing key inputs to the computer program.

Effective use of the basket method can give planners the flexibility to reprogram their limited analysis and negotiation resources in ways which may be more productive. For example, it is not uncommon to find that a multitude of low dollar proposals require about 70 percent of the time spent in analysis and negotiation but account for only 5 to 10 percent of the total backlog bid price. Assuming a 20 percent sampling rate is applied to the low dollar proposals (i.e., only one out of every 5 is actually analyzed and negotiated), a saving of 56 percent (four-fifths of 70 percent) of analysis and negotiation effort could be achieved. Resources thus saved could then be reprogrammed to deal with the large dollar proposals, resulting in near tripling the available manpower (from 30 percent to 30 + 56 = 86 percent). Alternately, by doubling the effort expended in analysis and negotiation of the sample of low dollar proposals, the savings are reduced from 56 to 42 percent (three-fifths of 70 percent) thus increasing resources available for the large dollar proposals from 30 percent to 30 + 42 = 72 percent. This translates to a 100 percent increase in available time for the low dollar proposals and a 140 percent increase for the large dollar proposals. While these figures will certainly differ from site-to-site, they give an indication of how the use of the basket method can increase management prerogatives in allocating scarce analysis and negotiation resources.

Part B

The "Basket Method": Basic Ideas

This part consists of two subsections:

1. How to select a "balanced" sample.
2. How to estimate negotiated prices for unsampled proposals.

Subsection 1. The name "Basket Method" derives from the manner in which the population is partitioned into separate groups prior to randomly selecting one of the groups as the sample. Imagine having a population of 100 proposals (N=100) from which a 10% sample (n=10) is to be selected. Assume the proposals have been arranged in order of decreasing bid price and indexed or numbered accordingly, that is, the proposal with the largest bid price in labled number 1, the next largest number 2, and so on. Our goal is to select a representative 10% sample. For now, we will think of "representative" only in terms of bid prices. Important other measures of representation are discussed in Part D.

By "representative", we will mean a sample of proposals whose bid prices look like those of the population, only thinned out a bit. There are over 17 trillion possible samples of size 10 that can be formed from 100 proposals. Some of those (e.g. the ten with smallest bid prices) would not be termed "representative". One way to avoid getting such an unbalanced sample would be to divide the population into 10 groups (1,11,21, ... 91), (2,12,22, ... 92), (3,13,23, ... 93), ... (10,20,30, ... 100) and select one group at random. This is called "systematic sampling". A second method, called "stratified sampling", selects one proposal at random from each of the groups (1,2,3, ... 10), (11,12,13, ... 20), (21,22,23, ... 30), ... (91,92,93 ... 100). While both of these standard methods (and many others) would result in a fairly representative sample, standard estimation rules based on such samples can be systematically exploited by the contractor since he has control on the state-of-nature.

For this reason, the Basket Method partitions the population into groups similar to those resulting from the systematic sampling scheme but does this in a way which eliminates the opportunity for gamesmanship such as the clever padding strategies discussed in reference (2). Using the previous example, imagine 10 baskets into which individual proposals will be placed. Starting with proposals 1 through 10, (those with the largest bid prices) place one proposal in each basket. Each successive group of 10 proposals are assigned, one-per-basket, using the following rule: the largest unassigned proposal is placed in the basket with the smallest sum of bid prices. For the second group of 10 proposals, this rule results in pairing 11 with 10, 12 with 9, ..., and 20 with 1. Basket subtotals are then calculated and the assignment rule applied to the third group of 10 proposals. This is repeated until all the proposals have been assigned. If N/n is not an integer, the last group (smallest bid prices) will be incomplete, resulting in some baskets being shorted one proposal. Due to the sequential balancing of basket totals at each stage of the process, this initial assignment should result in nearly equal basket totals. In the computer program which does the balancing, however, a swapping algorithm is used to bring basket totals into closer agreement after the initial assignments are made.

A simple example will illustrate the basket balancing process. Suppose N=8 and the bid prices are 79,76,61,54,39,34,24,10. Forming 2 baskets would go as follows. The first 2 groups are assigned in the obvious way resulting in:

Basket	Bid Prices	Subtotals
1	79,54	133
2	76,61	137

Since basket 1 now has the smaller subtotal, it receives the largest unassigned proposal (39) with the other member of the third group (34) being placed in basket 2 resulting in subtotals of 172 in basket 1 and 171 in basket 2. The largest unassigned proposal (24) goes to the basket with smaller subtotal (basket 2) and the other proposal (10) goes to basket 1. This results in an initial assignment of:

Basket	Bid Prices	Subtotals
1	79,54,39,10	182
2	76,61,34,24	195

The subtotals can be brought into closer agreement by swapping 61 for 54 resulting in:

Basket	Bid Prices	Totals
1	79,61,39,10	189
2	76,54,34,24	188

No additional swapping will improve the balance so the algorithm stops at this point. Note the "balance" or similarity between the two baskets. The average of the proposals in basket 1, basket 2, and the population are, respectively, 47.25, 47.00, and 47.125. Standard deviation are, respectively, 29.7, 23.0, and 24.6. Theoretical reasons for wanting as close a concordance as possible between basket and population statistics (means, standard deviations, etc.) are discussed in references (1) and (3). We refer to such a concordance as "balance".

By way of contrast, imbalance can be seen in the following examples. It can be shown that there are 70 different ways to place 8 proposals into 2 baskets of 4 each. Some of these will be fairly well balanced like the given example. Others may be terribly unbalanced in totals such as:

Basket	Bid Prices	Totals
1	79,76,61,54	270
2	39,34,24,10	107

or, balanced in totals but unbalanced in other ways such as:

Basket	Bid Prices	Totals
1	79,76,24,10	189
2	61,54,39,34	188

In this arrangement, basket 1 contains the 2 largest and 2 smallest proposals while basket 2 is composed of the intermediate values. The standard deviations, 35.4 and 12.6, respectively, signal this imbalance, differing by a factor of almost three. If the baskets are balanced, then no matter which basket is selected as the sample to be analyzed and negotiated, its composition of component bid prices will be representative of the composition of the remainder of the unsampled proposals thereby avoiding exploitation.

Balancing the baskets by hand is, if it can be done at all, quite tedious when the population is large and many baskets are used. To ease this burden, a computer program was developed and is available to users who are subscribers to the Air Force "COPPER IMPACT" computer services. Instructions for running this program are contained in the appendix.

Shown below is an example of the output of that program for the case of forming four baskets from a population of 120 proposals from an aerospace contractor.

BASKET NO. 1

	PROPOSAL NO.	BID PRICE
1.	104	89.495
2.	41	63.769
3.	106	60.488
4.	65	51.128
5.	56	42.272
6.	46	41.200
7.	112	35.618
8.	16	32.120
9.	75	31.294
10.	61	22.696
11.	114	20.864
12.	83	19.272
13.	71	18.344
14.	22	15.985
15.	101	13.728
16.	68	13.519
17.	38	10.346
18.	120	9.109
19.	59	9.003
20.	113	7.772
21.	77	6.188
22.	57	6.122
23.	15	4.710
24.	63	4.559
25.	82	4.094
26.	91	3.033
27.	117	2.411
28.	64	2.317
29.	86	1.952
30.	33	0.018

BASKET NO. 2

	PROPOSAL NO.	BID PRICE
1.	89	88.741
2.	18	64.171
3.	74	61.293
4.	19	49.803
5.	85	45.310
6.	118	39.977
7.	42	34.567
8.	2	33.983
9.	9	25.755
10.	28	25.667
11.	21	22.527
12.	26	19.253
13.	24	16.549
14.	92	16.546
15.	90	14.596
16.	87	11.797
17.	23	11.555
18.	10	10.126
19.	69	8.428
20.	54	6.884
21.	116	6.633
22.	49	5.416
23.	79	5.413
24.	55	4.144
25.	12	4.016
26.	44	3.073
27.	97	2.540
28.	48	2.313
29.	81	1.571
30.	6	0.785

BASKET NO. 3

	PROPOSAL NO.	BID PRICE
1.	100	87.522
2.	76	72.411
3.	32	57.736
4.	96	48.773
5.	4	41.378
6.	109	40.173
7.	31	37.552
8.	51	31.420
9.	94	28.733
10.	20	23.963
11.	11	22.195
12.	78	18.697
13.	29	17.197
14.	34	16.050
15.	35	14.133
16.	84	13.073
17.	103	11.334
18.	80	9.748
19.	40	8.402
20.	14	7.109
21.	107	6.781
22.	72	5.506
23.	95	5.243
24.	43	4.495
25.	5	3.825
26.	102	3.035
27.	66	2.684
28.	60	2.235
29.	52	1.507
30.	47	1.506

BASKET NO. 4

	PROPOSAL NO.	BID PRICE
1.	30	82.016
2.	13	60.874
3.	37	51.842
4.	99	48.931
5.	105	47.846
6.	62	39.784
7.	119	34.264
8.	58	31.398
9.	25	29.284
10.	73	23.777
11.	115	21.382
12.	70	19.258
13.	88	18.319
14.	8	15.798
15.	17	14.603
16.	39	12.192
17.	53	11.506
18.	27	9.106
19.	111	[illegible]
20.	98	[illegible]
21.	3	6.800
22.	108	5.984
23.	7	5.261
24.	93	4.117
25.	110	3.393
26.	45	3.386
27.	1	2.858
28.	50	2.124
29.	36	1.363
30.	67	0.380

BASKET STATISTICS

DIFFERENCE BETWEEN MAX AND MIN BASKETS = 0.016
MEAN ERROR BOUNDED BY-0.00% AND 0.00%.
BID PRICE MOMENTS:

BASKET	NO. PROPOSALS	TOTAL OF BIDS	MEAN	STD DEV
1	30	643.423	21.447	21.671
2	30	643.428	21.448	[illegible]
3	30	643.412	21.447	21.753
4	30	643.426	21.448	21.812
POPUL	120	2573.689	21.447	21.719

Note the excellent balance of bid prices among the baskets. A population in excess of $2.5 million was partitioned into 4 baskets with a maximum difference between basket totals of $16. Basket standard deviations are nearly identical. (Note: $ values in all printouts are expressed in thousands.)

Subsection 2. After the baskets are formed, one is selected at random. This can be accomplished in any convenient mutually agreeable way such as placing K slip of paper numbered 1 through K in a hat and selecting one at random or, if K is six or less, tossing a die. The proposals in the basket thus selected become the sample.

The proposals in the selected basket are then analyzed and negotiated in the normal manner. The resulting negotiated prices provide the basis for estimation.

The ESTIMATION RULE, called "ratio estimation", works as follows. Using the results of the sample negotiation, compute the sample ratio factor

$$R = \frac{\text{Total negotiated price of sample}}{\text{Total bid price of sample}}$$

and apply this factor to each unsampled proposal. If, for example, the total negotiated value were 85% of the total bid value, each unsampled proposal would have an estimated negotiated value of 85% of its bid value. There are several algebraically equivalent ways of viewing the ratio estimator:

a. The ratio of total negotiated to total bid price for the unsampled proposals is estimated to be the same as the ratio R computed from the sample.

b. The estimate of the total negotiated price T is given by

(Sum of negotiated values of sampled proposals) +
R×(Sum of bid prices of unsampled proposals)

c. If there are K baskets and the total bid price of each basket is the same, the estimate of T is given by

K×(Total negotiated price of sample)

When teamed with the Basket Method, the ratio estimator is invulnerable to gamesmanship. The estimates are unbiased, that is, they do not systematically favor either side. Using data from actual negotiations, experience has demonstrated a high degree of precision is attained using the Basket Method and ratio estimation.

While the computations required to estimate the prices of the unsampled proposals can readily be done with a small calculator, a computer program is provided on the COPPER IMPACT network to do these calculations and print out the results. Instructions for running that program and an example of the output are contained in the Part E.

Part C

Perspective Analysis: Tailoring the sampling procedure to local conditions.

The POPULATION, SAMPLING PLAN and ESTIMATION RULE were discussed in Part B. The fourth basic component common to all sampling and estimation problems, PROBABLE ERROR, will now be discussed. One cannot expect error-free inference based on partial information but most statistical procedures are accompanied by a statement concerning the PRECISION of the estimate. This may come in the form of something called the "standard error of the estimate" or perhaps a confidence interval. They are necessarily based on some sort of assumed probabilistic model for the relationship between X and Y. No such general model assumptions suggest themselves for the relationship between bid and negotiated prices, nor would it be appropriate to force this scenario into some inappropriate standard framework. Experience has shown that there are great differences between contractors in the manner in which bids are prepared and indeed, even between different divisions within a single contractor's organization. So, rather than obtaining the correct answer to the wrong problem by making doubtful model assumptions, our approach to sample size selection and a statement of probable error is based on a computerized perspective analysis of the historical relationship between bid and negotiated prices tailored to local conditions. Its validity is based on the assumption that statistical patterns in the relationship between bid and negotiated prices which existed in the recent past will tend to persist in the near future. This assumption seems reasonably justified if

a. directives governing preparation, analysis, and negotiation of price proposals are stable,

b. analysis and negotiation practices of on-site government personnel conform to these directives, and

c. the contractor's estimation policies and personnel have not substantially changed over the period of time involved.

These conditions tend to result in a statistical stability which can be exploited in determining an appropriate sample size by using the perspective analysis program. Because the Basket Method will be employed, the sample size must necessarily be a simple fractional part of the population size. Nevertheless, forming K baskets and picking one at random gives a sufficiently "rich" set of alternative sample sizes for most applications. Which sample size is appropriate for a given application requires a decision based on local conditions. PERSPECTIVE ANALYSIS provides a rational basis for the sample size decision as will be seen.

To use the PERSPECTIVE ANALYSIS program the user assembles a collection of recently negotiated proposals which are typical of the type of proposals expected to constitute the population on which actual sampling is contemplated in the future. As a rule of thumb, this historical "superpopulation" should be at least a fourth again as numerous as anticipated population sizes. If one anticipated populations of around N=100, the historical data set should contain upwards of 125 negotiated proposals with the larger-the-better up to the program storage limit currently set at 350. A file containing these bid and negotiated prices is created as explained in Part E. In response to commands issued by the user, the perspective analysis program simulates what <u>would</u> have happened had sampling and estimation been used to price out different propulations formed from the superpopulation. With the historical superpopulation serving as a "generator" of typical populations, a large number of such populations is produced. Each time a population is formed it is partitioned into K baskets (K being specified by the user), one of which is selected at random. Since bid and negotiated prices of the proposals in this selected basket are known, the sample ratio factor

$$R = \frac{\text{total negotiated price of proposals in selected basket}}{\text{total bid price of proposals in selected basket}}$$

can be computed and applied to estimate the total negotiated price in the population created. Since the actual total negotiated value of this population is known, a percent error due to sampling is calculated:

$$\%\ \text{error} = \frac{100\ (\text{estimated negotiation total} - \text{actual negotiation total})}{(\text{actual negotiation total})}$$

This error may be positive (overaward) or negative (underaward). These percentages are then tallied into "class intervals" of width 1%, class interval frequencies are tabulated and "bar charts" or "histograms" are produced. A set of such histograms will now be displayed for data reflecting actual bid and negotiated prices of 120 proposals from an aerospace contractor. Output is shown for 2,3,4,5 and 6 basket setups based on 500 replications of the process described above.

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 2 BASKETS
BAR CHART

CLASS INTERVAL CENTER	CUM REL FREQ	REL FREQ	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.0	0.0	I
-8.00	0.0	0.0	I
-7.00	0.0	0.0	I
-6.00	0.0	0.0	I
-5.00	0.0	0.0	I
-4.00	0.0	0.0	I
-3.00	0.0	0.0	I
-2.00	0.024	0.024	I**
-1.00	0.284	0.260	I*************
0.0	0.712	0.428	I**********************
1.00	0.976	0.264	I**************
2.00	1.000	0.024	I**
3.00	1.000	0.0	I
4.00	1.000	0.0	I
5.00	1.000	0.0	I
6.00	1.000	0.0	I
7.00	1.000	0.0	I
8.00	1.000	0.0	I
9.00	1.000	0.0	I
10.00	1.000	0.0	I

PERCENT ERROR: MEAN = -0.001 STD DEV = 0.800

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 3 BASKETS
BAR CHART

CLASS INTERVAL CENTER	CUM REL FREQ	REL FREQ	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.0	0.0	I
-8.00	0.0	0.0	I
-7.00	0.0	0.0	I
-6.00	0.0	0.0	I
-5.00	0.0	0.0	I
-4.00	0.0	0.0	I
-3.00	0.020	0.020	I*
-2.00	0.096	0.076	I****
-1.00	0.308	0.212	I***********
0.0	0.664	0.356	I******************
1.00	0.908	0.244	I*************
2.00	0.994	0.086	I*****
3.00	1.000	0.006	I*
4.00	1.000	0.0	I
5.00	1.000	0.0	I
6.00	1.000	0.0	I
7.00	1.000	0.0	I
8.00	1.000	0.0	I
9.00	1.000	0.0	I
10.00	1.000	0.0	I

PERCENT ERROR: MEAN = 0.022 STD DEV = 1.122

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 4 BASKETS
BAR CHART

CLASS INTERVAL CENTER	CUM REL FREQ	REL FREQ	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.0	0.0	I
-8.00	0.0	0.0	I
-7.00	0.0	0.0	I
-6.00	0.0	0.0	I
-5.00	0.0	0.0	I
-4.00	0.004	0.004	I*
-3.00	0.020	0.016	I*
-2.00	0.100	0.080	I****
-1.00	0.310	0.210	I***********
0.0	0.622	0.312	I****************
1.00	0.856	0.234	I************
2.00	0.966	0.110	I******
3.00	0.994	0.028	I**
4.00	1.000	0.006	I*
5.00	1.000	0.0	I
6.00	1.000	0.0	I
7.00	1.000	0.0	I
8.00	1.000	0.0	I
9.00	1.000	0.0	I
10.00	1.000	0.0	I

PERCENT ERROR: MEAN = 0.127 STD DEV = 1.289

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 5 BASKETS
BAR CHART

CLASS INTERVAL CENTER	CUM REL FREQ	REL FREQ	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.0	0.0	I
-8.00	0.0	0.0	I
-7.00	0.0	0.0	I
-6.00	0.0	0.0	I
-5.00	0.002	0.002	I*
-4.00	0.032	0.030	I**
-3.00	0.082	0.050	I***
-2.00	0.190	0.108	I******
-1.00	0.392	0.202	I***********
0.0	0.622	0.230	I************
1.00	0.814	0.192	I**********
2.00	0.926	0.112	I******
3.00	0.990	0.064	I****
4.00	1.000	0.010	I*
5.00	1.000	0.0	I
6.00	1.000	0.0	I
7.00	1.000	0.0	I
8.00	1.000	0.0	I
9.00	1.000	0.0	I
10.00	1.000	0.0	I

PERCENT ERROR: MEAN = -0.047 STD DEV = 1.698

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 6 BASKETS
BAR CHART

CLASS INTERVAL CENTER	CUM REL FREQ	REL FREQ	0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.0	0.0	I
-8.00	0.0	0.0	I
-7.00	0.0	0.0	I
-6.00	0.002	0.002	I*
-5.00	0.006	0.004	I*
-4.00	0.036	0.030	I**
-3.00	0.090	0.054	I***
-2.00	0.198	0.108	I******
-1.00	0.420	0.222	I************
0.0	0.638	0.218	I***********
1.00	0.832	0.196	I**********
2.00	0.938	0.106	I******
3.00	0.978	0.040	I**
4.00	0.996	0.018	I*
5.00	1.000	0.004	I*
6.00	1.000	0.0	I
7.00	1.000	0.0	I
8.00	1.000	0.0	I
9.00	1.000	0.0	I
10.00	1.000	0.0	I

PERCENT ERROR: MEAN = -0.129 STD DEV = 1.757

Examining the first of these (2 baskets or a 50% sample), note the first column of numbers are class interval centers ranging from -10% to +10%. (Had any errors been larger than ±10%, they would have been separately listed.) The third column of figures is the relative frequency of occurrences of errors within ±1/2% of the class interval center. For example, looking at the line labled 1.00, we find that a fraction 0.264 of the time (i.e., 132 out of 500 runs) sampling resulted in an overaward of between 0.5% and 1.5%. This 0.264 frequency is represented by the corresponding bar length +++++++++++++* where each "+" has a value of .02 and the "*" represents any value less than a full .02 increment. The second column of numbers labled "CUM REL FREQ" is the cumulative relative frequency up to that point. Again, looking at the 1.00 row, we find the entry 0.976 which indicates that 97.6% of the sampling errors fell in some class interval up to and including 1.00%. This relative error frequency data is also summarized in a table of <u>central</u> error frequencies. Examining this table we see, for example, that 95.8% of the errors fell in class intervals centered between -3% and +3% using 20 percent sampling (i.e., five baskets).

Number of Baskets	Central Error Frequencies 0	±1%	±2%	±3%	±4%
2	.428	.952	1.000	1.000	1.000
3	.356	.812	.974	1.000	1.000
4	.312	.756	.946	.990	1.000
5	.230	.624	.844	.958	.998
6	.216	.634	.848	.942	.990

Based on the histograms, central error frequencies, and other considerations of local importance, the decision maker chooses the number of baskets to be used in actual application. No hard-and-fast rules can be given for this administrative choice. As the number of baskets increases, the sample size decreases and, as is always the case with smaller samples, the opportunity for estimation error increases. The word "opportunity" should be stressed. In any given application, we cannot say in advance just what the error due to sampling will be. In repeated applications, however, these histograms are indicative of the long term behavior of such errors. In this respect, there is neither a systematic tendency to overaward or underaward due to sampling and, on the average, the errors will cancel out over the long run. This is true whether 2 baskets (50% sample) or 10 baskets (10% sample) are used but 2 baskets give better protection against making an error as large as, say, +4% than does 10 baskets in any single sample. Of course, 2 baskets require 5 times as much analysis and negotiation as does 10 baskets. Hence the decision is a trade-off between efficiency and accuracy (error opportunity).

In Part D on fine tuning, the topic of perspective analysis will be explored in further detail. There it will be shown how balancing on other relevant characteristics (in addition to bid price) can increase precision for a given sample size or, equivalently, reduce sample size requirements for a given level of precision.

Part D

Fine Tuning: Optional Use of Multilevel Attributes

The merit of a sample being "balanced" on bid price structure is appealing because if the baskets are similar in bid price composition there should be no strong apriori reason to believe that one basket should have a higher negotiated total than any other basket. And if the negotiated totals were the same for each basket, then its easy to see that no error would be made by negotiating a single basket and multiplying its total negotiated value by the number of baskets in order to estimate the total negotiated value of all proposals in the population.

But, besides bid price, there may be other relevant factors that should be considered in our definition of a "balanced" sample. For example, suppose basket 1 ended up being composed of predominately "high technology" proposals while the rest of the baskets were composed of predominately "off-the-shelf" type items. Suppose further that the contractor's estimation group tended to add excessive "contingency factors" to high technology jobs to allow leeway for unforseen problems while the off-the-shelf jobs were usually negotiated at or near the bid price. If there were K baskets in all, then there would be one chance out of K (basket 1 is chosen for the sample) for a large underaward since the decrement in basket 1 would be wrongly applied to the remaining (K-1) baskets. On the other hand, if any basket besides basket 1 is chosen, and the chances are (K-1) out of K for this to happen, there would be a smaller overaward since basket 1 would not be properly decremented. While undesirable, this type of imbalance does not alter the unbiasedness of the basket method. It can be shown, even in cases such as this, that overawards and underawards tend to cancel out so that, in the long run, the average overaward is zero.

This type of imbalance does have an adverse effect on sampling error in terms of decreased precision, however. By taking additional relevant characteristics into consideration when the baskets are balanced, precision can often be dramatically increased. Guidelines on selecting relevant characteristics and obtaining samples which are balanced on these characteristics are briefly discussed in this section.

Unfortunately, no sure-thing rules can be given for selecting good characteristics. Even the definition of "good" is difficult to nail down. What is important, of course, is the effect that balancing with respect to a particular characteristic has on the percent error histograms. This can be assessed empirically by running the perspective analysis simulation with and without balancing on the given characteristic.

A MULTILEVEL CHARACTERISTIC is any qualitative feature which partitions proposals into separate groups such as:

a. Work area (hull, propulsion, electrical, etc.)

b. Type of work (change order, repair, retrofit kit, etc.)

c. Degree of labor intensity (high, low)

d. Level of technology required (advanced, moderate, basic)

"a" and "b" above are examples of pure qualitative characteristics while "c" and "d" are quantitative characteristics which have been made qualitative by categorization. "c" might be defined by computing the ratio of direct labor dollars to total direct cost with "low" meaning 20% or less, say. One could define a 6-level characteristic by pairing the levels of "c" with those of "d" obtaining jobs sorted by technology (T) and intensity of labor (L), that is: advanced T, high L; advanced T, low L; moderate T, high L; moderate T, low L; basic T, high L; and basic T, low L. Only imagination limits the kind of characteristics which can be tried but judgment should be exercised in selecting relevant characteristics since considerable computer time may be wasted otherwise.

In selecting a multilevel characteristic to try, the user should exploit local analysis and negotiation experience. He should look for a characteristic which tends to partition proposals into groups which are more homogeneous within themselves with respect to percent decrement than the population taken as a whole. Percent decrement is defined as 100×(dollar decrement/bid price). Ideally (and this is only a target, not a realizable goal in practice) the groups formed by the characteristic should have a very narrow range of percent decrement such as, say:

Group 1	over 15%
Group 2	10 - 15%
Group 3	4 - 10%
Group 4	under 4%

In advance of actual analysis and negotiation, one can never be <u>certain</u> that a proposal of a particular type (e.g., one involving a state-of-the-art high technology fabrication) will, in fact, have a large contingency pad, but experience with proposals of a similar nature may suggest a <u>tendency</u> in that direction. Similarly, off-the-shelf low-labor assemblies might tend to be negotiated at or near the bid price.

Exploring historical data in search of relevant characteristics which will reduce sampling error can involve extensive use of sophisticated statistical techniques. Experience indicates, however, that if there <u>are</u> any characteristics which can be discovered through such data analyses, those characteristics are the ones which would be identified as potentially important by experienced negotiators. Typically, these characteristics differ from site-to-site simply because of differences in the way individual contractors conduct their estimation processes. On-site government analysts know from experience how the local process operates and are best equipped to identify characteristics which should be taken into account in balancing baskets. Indeed, the best characteristics may be subjectively defined as was the case in Example 2 which follows.

Balancing on a multilevel characteristic may not be necessary if the error historgrams are satisfactory without this fine tuning. There are also situations where no characteristics can be found which appreciably increase estimation precision. But the cost of looking is small compared with the dramatic improvements which have been obtained in some cases. The same programs are used whether or not proposal characteristics are included in the analysis. Differences occur in the way data files are structured to handle the extra information and in the way various prompts are answered as shown in the appendix. We illustrate the use of these programs with two examples:

Example 1. The historical data base consists of N = 120 proposals from a shipbuilding activity. Unlike the next example, the data are fairly "well behaved" with percent decrements rather evenly spread over the range from 26% down to -2% (a negative decrement is an increment). Bid prices range from $98,557 down to $2,280. Three multilevel factors were suggested as possibly relevant:

Factor	Levels	Level Names
Job Area	3	Hull, Machinery, Electrical
Ship Type	2	Carriers, Others
Labor Intensity	2	High, Low

These factors can be tried separately or together in a variety of combinations. For example, one could label the proposals with a "1" or "2" depending on ship type only, with "1" standing for proposals relating to work done on carriers. Or, one could use a cross-classification using Job Area and Labor Intensity with six combinations: hull, high (1); hull, low (2); machinery, high (3); machinery, low (4); electrical, high (5); and electrical, low (6). By combining all levels of the three factors in the obvious way, one could come up with a classification scheme with 12 characteristics. One could experiment with all possible combinations of these factors using the perspective analysis simulation program but the time, effort, and computation expense for this kind of exhaustive approach can be avoided by applying some statistical analysis (3-way Analysis of Variance) or, more simply, by looking at a few numbers and making educated guesses. If one calculates the percent decrement for each proposal, averages these for proposals in each of the 12 groups and tabulates the results in some covenient form like:

	Carriers		Other Ship Construction	
	High Labor	Low Labor	High Labor	Low Labor
Hull	8.8	3.4	9.6	3.0
Machinery	9.1	3.8	8.9	3.2
Electrical	15.1	10.7	14.7	11.3

one observes a high degree of similarity between the average percent decrement by column of the six numbers under "carriers" and the corresponding six numbers under "Other Ship Construction". This suggests that distinguishing between carriers and other ships is not going to make any difference. Combining these two groups and reaveraging the proposals in the composite categories results in a new table:

	High Labor	Low Labor
Hull	9.2	3.2
Machinery	9.0	3.5
Electrical	14.9	11.0

Again, we observe little difference between the first 2 rows of this table but a large difference between these and the third row. This suggests that Hull and Machinery could be lumped together. Based on these observations a 4-type characteristic suggested itself: Hull/Mach, High (1); Hull/Mach, Low (2); Elect, High (3); and Elect, Low (4).

It is useful to run the simulation program twice: once without and once with the characteristics. This provides the opportunity to compare error histograms and learn what, if any, is the effect of including characteristics other than bid price alone. To save space, only the 4 basket histograms are displayed but we include the central error frequencies for 2,3, and 4 baskets for comparison purposes.

```
                         SIMULATION RESULTS:
                 RELATIVE FREQUENCY OF PERCENT OVERAWARD
                          WITH  4 BASKETS
 CLASS      CUM                BAR CHART
INTRVAL     REL       REL
CENTER      FREQ      FREQ       0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
-----------------------------I----I----I----I----I----I----I----I----I----I-----
 -10.00    0.0       0.0     I
  -9.00    0.0       0.0     I
  -8.00    0.0       0.0     I
  -7.00    0.0       0.0     I
  -6.00    0.0       0.0     I
  -5.00    0.0       0.0     I
  -4.00    0.001     0.001   I*
  -3.00    0.021     0.020   I+
  -2.00    0.109     0.088   I++++*
  -1.00    0.349     0.240   I+++++++++++*
   0.0     0.661     0.312   I+++++++++++++++*
   1.00    0.891     0.230   I+++++++++++*
   2.00    0.970     0.079   I+++*
   3.00    0.996     0.026   I+*
   4.00    1.000     0.004   I*
   5.00    1.000     0.0     I
   6.00    1.000     0.0     I
   7.00    1.000     0.0     I
   8.00    1.000     0.0     I
   9.00    1.000     0.0     I
  10.00    1.000     0.0     I
-----------------------------I----I----I----I----I----I----I----I----I----I-----
 PERCENT ERROR:    MEAN =  -0.002   STD DEV =    1.241
```

without using the 4 characteristics

```
                         SIMULATION RESULTS:
                 RELATIVE FREQUENCY OF PERCENT OVERAWARD
                          WITH  4 BASKETS
 CLASS      CUM                BAR CHART
INTRVAL     REL       REL
CENTER      FREQ      FREQ       0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
-----------------------------I----I----I----I----I----I----I----I----I----I-----
 -10.00    0.0       0.0     I
  -9.00    0.0       0.0     I
  -8.00    0.0       0.0     I
  -7.00    0.0       0.0     I
  -6.00    0.0       0.0     I
  -5.00    0.0       0.0     I
  -4.00    0.0       0.0     I
  -3.00    0.002     0.002   I*
  -2.00    0.066     0.064   I+++*
  -1.00    0.311     0.245   I++++++++++++*
   0.0     0.698     0.387   I++++++++++++++++++++*
   1.00    0.940     0.242   I++++++++++++*
   2.00    0.995     0.055   I++*
   3.00    1.000     0.005   I*
   4.00    1.000     0.0     I
   5.00    1.000     0.0     I
   6.00    1.000     0.0     I
   7.00    1.000     0.0     I
   8.00    1.000     0.0     I
   9.00    1.000     0.0     I
  10.00    1.000     0.0     I
-----------------------------I----I----I----I----I----I----I----I----I----I-----
 PERCENT ERROR:    MEAN =  -0.014   STD DEV =    0.977
```

using the 4 characteristics

Number of Baskets	Central Error Frequencies*			
	0	±1%	±2%	±3%
2	.523(.608)	.961(.998)	.999(1.000)	1.000(1.000)
3	.360(.511)	.826(.959)	.979(.999)	1.000(1.000)
4	.312(.387)	.782(.874)	.949(.993)	.995(1.000)

*Note: Entries in parentheses are frequencies for the runs which used the 4-level chanracteristic.

By inspecting the histograms themselves and the central error frequencies, the effect of using the characteristic can be seen. Because this data was so "well behaved" in the first place, the modest increase in precision may not justify the extra trouble of classifying proposals. Incidentally, when a run using all 12 combinations of the 3 factors was tried, the results were slightly worse than in the above simulation. Too many categories makes it difficult for the balancing algorithm to achieve overall balance because of the number of simultaneous constraints trying to be met. The present program is in fact limited to a maximum of 12 categories for that reason. Experience with real data suggests that use of more than 4 or 5 categories is rarely justified.

Example 2: The scenario for this application differed in many substantive ways from the standard pricing situation for which the basket method was designed. Firstly, the proposals were from different corporations. Each proposal was prepared by a candidate subcontractor and pertained to components of a large follow-on procurement of a major weapons system. Secondly, the results of sampling, negotiation, and estimation were used to set a total target price to be used in an incentive agreement with the prime contractor. Thirdly, the proposals involved ranged in price from $100,000 to over $1,000,000 whereas interpretations of the DAR had previously limited the size of proposals which formed the population to under $100,000.

The procedure for the subcontracts portion of the annual follow-on buy starts with the prime contractor assembling subcontractor bids. After some preliminary analysis on his part, he submits a proposed target price for the entire package. This encompasses in the neighborhood of 80 component proposals totaling about 100 million dollars. The cognizant government agency then marshals a task force to visit each subcontractor to analyze the various cost elements, check prices and quantities, audit appropriate records, etc. This is an involved process which puts considerable time demands on limited personnel resources. It is generally felt that the heavy workload works against thoroughness and accuracy. Based on these analyses, the agency establishes its target price. The difference between this value and the prime contractor's proposed total is then negotiated to arrive at a firm target price. The prime contractor then negotiates individual prices with each subcontractor and receives a portion of any decrement below the total target price he is able to negotiate under the provisions of an incentive clause in the prime contract.

The cognizant agency felt it could produce a more realistic and firm target price if its analysis resources could be concentrated on a few subcontractors rather than being spread thinly over the entire group. To check out the feasibility of basing a target price on an analyzing sample of subcontractor proposals, historical data were assembled on bid and negotiated prices and the basket method simulation program was run without characteristics. The percent error histograms and central error frequencies are shown below.

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 2 BASKETS

CLASS INTRVAL CENTER	CUM REL FREQ	REL FREQ	BAR CHART 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.0	0.0	I
-8.00	0.0	0.0	I
-7.00	0.0	0.0	I
-6.00	0.001	0.001	I*
-5.00	0.008	0.007	I*
-4.00	0.035	0.027	I+*
-3.00	0.113	0.078	I+++*
-2.00	0.237	0.124	I++++++*
-1.00	0.414	0.177	I++++++++*
0.0	0.601	0.187	I+++++++++*
1.00	0.754	0.153	I+++++++*
2.00	0.893	0.139	I++++++*
3.00	0.966	0.073	I+++*
4.00	0.989	0.023	I+*
5.00	0.998	0.009	I*
6.00	1.000	0.002	I*
7.00	1.000	0.0	I
8.00	1.000	0.0	I
9.00	1.000	0.0	I
10.00	1.000	0.0	I

PERCENT ERROR: MEAN = -0.007 STD DEV = 1.992

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 3 BASKETS

CLASS INTRVAL CENTER	CUM REL FREQ	REL FREQ	BAR CHART 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.0	0.0	I
-8.00	0.0	0.0	I
-7.00	0.004	0.004	I*
-6.00	0.012	0.008	I*
-5.00	0.047	0.035	I+*
-4.00	0.101	0.054	I++*
-3.00	0.168	0.067	I+++*
-2.00	0.292	0.124	I++++++*
-1.00	0.434	0.142	I+++++++*
0.0	0.575	0.141	I+++++++*
1.00	0.719	0.144	I+++++++*
2.00	0.840	0.121	I++++++*
3.00	0.917	0.077	I+++*
4.00	0.961	0.044	I++*
5.00	0.983	0.022	I+*
6.00	0.995	0.012	I*
7.00	0.999	0.004	I*
8.00	1.000	0.001	I*
9.00	1.000	0.0	I
10.00	1.000	0.0	I

PERCENT ERROR: MEAN = -0.048 STD DEV = 2.580

SIMULATION RESULTS:
RELATIVE FREQUENCY OF PERCENT OVERAWARD
WITH 4 BASKETS

CLASS INTRVAL CENTER	CUM REL FREQ	REL FREQ	BAR CHART 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
-10.00	0.0	0.0	I
-9.00	0.001	0.001	I*
-8.00	0.003	0.002	I*
-7.00	0.011	0.008	I*
-6.00	0.040	0.029	I+*
-5.00	0.090	0.050	I++*
-4.00	0.174	0.084	I++++*
-3.00	0.255	0.081	I++++*
-2.00	0.347	0.092	I++++*
-1.00	0.439	0.092	I++++*
0.0	0.567	0.128	I++++++*
1.00	0.675	0.108	I+++++*
2.00	0.774	0.099	I++++*
3.00	0.853	0.079	I+++*
4.00	0.898	0.045	I++*
5.00	0.935	0.037	I+*
6.00	0.964	0.029	I+*
7.00	0.979	0.015	I*
8.00	0.994	0.015	I*
9.00	0.999	0.005	I*
10.00	1.000	0.001	I*

PERCENT ERROR: MEAN = 0.013 STD DEV = 3.394

Number of Baskets	Central Error Frequencies 0	±1%	±2%	±3%	±4%
2	.187	.517	.780	.931	.981
3	.141	.427	.672	.816	.914
4	.128	.328	.519	.679	.808

Based on these results, the risk of large error was deemed unacceptable. In an attempt to explore for ways to improve these results, the prime contractor was asked to partition the subcontractors into 3 groups based on past experience relating to the quality of the subcontractor's estimation organization and policies concerning preparation of bids. The prime contractor, based on subjective assessment, characterized each subcontractor as good, fair or poor. These characteristics were included in the data and a new set of simulation runs was conducted resulting in the following error histograms and central error frequencies.

```
                              SIMULATION RESULTS:
                    RELATIVE FREQUENCY OF PERCENT OVERAWARD
                                WITH  2 BASKETS
 CLASS        CUM                      BAR CHART
INTRVAL       REL         REL
CENTER        FREQ        FREQ          0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
----------------------------------I----I----I----I----I----I----I----I----I----I----

 -10.00      0.0          0.0       I
  -9.00      0.0          0.0       I
  -8.00      0.0          0.0       I
  -7.00      0.0          0.0       I
  -6.00      0.0          0.0       I
  -5.00      0.0          0.0       I
  -4.00      0.0          0.0       I
  -3.00      0.006        0.006     I*
  -2.00      0.059        0.053     I++*
  -1.00      0.282        0.223     I++++++++++++*
   0.0       0.721        0.439     I+++++++++++++++++++++++*
   1.00      0.949        0.228     I++++++++++++*
   2.00      0.997        0.048     I++*
   3.00      1.000        0.003     I*
   4.00      1.000        0.0       I
   5.00      1.000        0.0       I
   6.00      1.000        0.0       I
   7.00      1.000        0.0       I
   8.00      1.000        0.0       I
   9.00      1.000        0.0       I
  10.00      1.000        0.0       I
----------------------------------I----I----I----I----I----I----I----I----I----I----

  PERCENT ERROR:   MEAN =  -0.008    STD DEV =    0.912
```

```
                              SIMULATION RESULTS:
                    RELATIVE FREQUENCY OF PERCENT OVERAWARD
                                WITH  3 BASKETS
 CLASS        CUM                      BAR CHART
INTRVAL       REL         REL
CENTER        FREQ        FREQ          0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
----------------------------------I----I----I----I----I----I----I----I----I----I----

 -10.00      0.0          0.0       I
  -9.00      0.0          0.0       I
  -8.00      0.0          0.0       I
  -7.00      0.0          0.0       I
  -6.00      0.0          0.0       I
  -5.00      0.0          0.0       I
  -4.00      0.002        0.002     I*
  -3.00      0.027        0.025     I+*
  -2.00      0.144        0.117     I+++++*
  -1.00      0.362        0.218     I+++++++++++*
   0.0       0.667        0.305     I++++++++++++++++*
   1.00      0.880        0.213     I+++++++++++*
   2.00      0.972        0.092     I++++*
   3.00      0.999        0.027     I+*
   4.00      1.000        0.001     I*
   5.00      1.000        0.0       I
   6.00      1.000        0.0       I
   7.00      1.000        0.0       I
   8.00      1.000        0.0       I
   9.00      1.000        0.0       I
  10.00      1.000        0.0       I
----------------------------------I----I----I----I----I----I----I----I----I----I----

  PERCENT ERROR:   MEAN =  -0.050    STD DEV =    1.297
```

```
                              SIMULATION RESULTS:
                    RELATIVE FREQUENCY OF PERCENT OVERAWARD
                                WITH  4 BASKETS
 CLASS        CUM                      BAR CHART
INTRVAL       REL         REL
CENTER        FREQ        FREQ          0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8  0.9
----------------------------------I----I----I----I----I----I----I----I----I----I----

 -10.00      0.0          0.0       I
  -9.00      0.0          0.0 ,     I
  -8.00      0.0          0.0       I
  -7.00      0.0          0.0       I
  -6.00      0.0          0.0       I
  -5.00      0.0          0.0       I
  -4.00      0.005        0.005     I*
  -3.00      0.051        0.046     I++*
  -2.00      0.187        0.136     I+++++++*
  -1.00      0.393        0.206     I++++++++++++*
   0.0       0.636        0.243     I++++++++++++++*
   1.00      0.822        0.186     I+++++++++++*
   2.00      0.933        0.111     I++++++*
   3.00      0.978        0.045     I++*
   4.00      0.998        0.020     I+
   5.00      1.000        0.002     I*
   6.00      1.000        0.0       I
   7.00      1.000        0.0       I
   8.00      1.000        0.0       I
   9.00      1.000        0.0       I
  10.00      1.000        0.0       I
----------------------------------I----I----I----I----I----I----I----I----I----I----

  PERCENT ERROR:   MEAN =   0.001    STD DEV =    1.610
```

Number of Baskets	Central Error Frequencies 0	±1%	±2%	±3%	±4%
2	.439	.890	.991	1.000	1.000
3	.305	.736	.945	.997	1.000
4	.243	.635	.882	.973	.998

Based on these results, an agreement to employ 4 baskets was reached. Both parties to the agreement are satisfied that equitable target prices have resulted from the first 2 applications and a third application is underway.

In summarizing this section, perspective analysis is a tool which allows potential users to examine in advance the effects of employing the basket method to price proposal backlogs. The results can be used to tailor the sample size to meet local conditions.

Part E

Guidelines and Examples

By this point in the exposition, the reader should have a good idea about what a balanced sample is and how the basket method tries to achieve this degree of representation. To supplement the general guidelines discussed throughout the first four parts of this document, this section provides some additional ideas, suggestions and examples intended to aid users in implementing the basket method. Detailed instructions for creating data files and using the basket method programs available through the COPPER IMPACT computer network are contained in the appendix.

The computer should not be expected, nor is it able, to replace good judgment by experienced users. It does what it's programmed to do, that is, manipulate the numbers it is fed, perform error free arithmetic and produce readable output. Intelligence is supplied by the user who assembles the input and must carefully study the output from the computer. Together, the user and the computer perform complementary functions.

When assembling the population from which the sample will be selected, the user should keep the goal of homogeniety in mind. For example, suppose 4 baskets are to be formed from 80 proposals. Suppose further that 3 of the proposals are in the 90-95K dollar range and the remaining ones are each less than 40K bid. There is obviously no way to achieve good balance in this case. Common sense suggests "scrubbing" the population of the three relatively large proposals and applying the basket method to the remaining 77 proposals under 40K. The three larger proposals would be negotiated separately, having no relation or influence on the sampling and estimation process. In a similar vein, suppose a 3 level characteristic were being used on a regular basis. A population of 66 proposals is assembled and examined. It is found that there are only 4 "type 1" proposals in this population with bid prices of 90K, 60K, 10K and 3K dollars, respectively. There is obviously no way to balance the "type 1" proposals among any number of baskets. It would be better to exclude these 4 proposals from the population and use a bilevel characteristic for this particular population. Anomalies like these should be detected when the population is being prepared for basketing but, in case of an oversight here, careful examination of the formed baskets will reveal the problem. While it is impossible to anticipate all such situations, careful inspection and application of good judgment should eliminate most problems.

While scrubbing the population, each proposal should receive a "desk audit" to insure that quantities are correct, rates are current, arithmetic is properly performed, etc., since once the sample is selected changes such as these contribute to the difference between bid and negotiated price. If arithmetic errors are found among the sampled proposals, it is reasonable to expect them to occur in the unsampled portion as well, since the sample is representative.

An effort has been made in this document to provide potential users with complete information on how the basket sampling method works. Other documentation cited in the bibliography deal with the basic theory, references (1) and (3), and provide a listing of internally documented computer codes, reference (4). These sampling techniques have been successfully employed in a wide variety of situations with excellent results. It is anticipated that formal arrangements will be established to assist field activities in obtaining assistance in the form of technical briefings and consultation, and that a user's share-group will be established to exchange applications information.

Part F

Appendix: Use of computer programs

The appendix contains detailed instructions on use of the Basket Method programs available through COPPER IMPACT, an interactive time-sharing computer network administered by the Air Force. The present COPPER IMPACT contract is with the General Electric Information Services Company which provides worldwide access through their "Mark III Service." DOD activities may obtain information from the Air Force Systems Pricing Office, Andrews Air Force Base (301) 981-4008 or AV 858-4008. We will not attempt to duplicate general information (e.g., logon procedures, editing, etc.) contained in the G. E. Command System Reference Manual which is provided to all system subscribers except to show those special commands required to run Basket Method programs.

All the Basket Method programs are accessed by issuing the single command "run basket*". The computer then prompts the user to identify which of three programs will be run:

(1) "Basket Simulation Model". This is the program used in perspective analysis of historical bid and negotiated prices as discussed in Parts C and D.

(2) "Basket Selection Model". This is the program used to form balanced baskets as discussed in Part B.

(3) "Award Model". This is the program used to calculate estimated prices of unsampled proposals based on negotiated prices of the sampled proposals mentioned in Part B.

In what follows, we give examples of each of these programs together with instruction for creating input data files.

To create an input data file prior to envoking any one of the Basket Method programs, the user informs the system that a new file is about to be created. This is accomplished with the command "new" followed by the name of the file to be created. This name, chosen by the user, can be any convenient name with 8 or fewer characters starting with an alphabetic character. Data is then entered in the "Data Entry Mode" (DEM) and saved as described in the Command System Reference Manual.

Example (1): "Basket Simulation Model".

Prior to running this model, the user must create a file of historical bid and negotiated prices for use as input to the program which performs perspective analysis. The format for a single line of input for this type of data set consists of a bid price, a negotiated price (each expressed in $1000's) and, if used, an integer identifying the characteristic type of the proposal. Entries are separated by commas. Assume the user selects the name "histdata" for this file. After the logon process is completed and the computer terminal is in the "READY" mode, commands are issued and data entered as shown below (uppercase lines are prompts issued by the computer). Each input line is ended with a carriage return and the DEM mode is terminated by depressing the "break" key. Dots (...) have been inserted for brevity instead of showing the entire input.

```
READY
new histdata

READY
dem

READY FOR INPUT
452.5,437,1
887,789,3
   .
   .
   .
304,293.5,2
609,542,3

READY
save

READY
```

Having created this file of historical data, the simulation program discussed in Parts C and D is run as shown below. The user is prompted for input with a question mark (?) following a brief description of what input is needed.

The data used came from the application discussed in Example 2 of Part D in which a subjective assessment of proposal quality was used to define a three-level characteristic: good, fair, and poor. The histograms shown in that section resulted from the following session on the computer.

```
READY
run basket*

BASKET*     12:27PST    12/11/81

WHICH OF THE FOLLOWING WOULD YOU LIKE TO RUN:

  1) BASKET SIMULATION MODEL
  2) BASKET SELECTION MODEL
  3) AWARD MODEL
  4) TERMINATE

INPUT 1,2, OR 3
 ?1

****************************************************************************
*                                                                          *
*                         PERSPECTIVE ANALYSIS OF                          *
*                              BASKET METHOD                               *
*            APPLIED TO HISTORICAL BID/NEGOTIATED RELATIONSHIPS            *
*                                                                          *
*                         SIMULATION AND PRICING                           *
*                         VERSION IV - JULY 1981                           *
*                                                                          *
*              THEORETICAL DEVELOPMENT - DR. K. T. WALLENIUS               *
*          ALGORITHM DESIGN AND IMPLEMENTATION - DR. STEPHEN BENZ          *
*                                                                          *
*                            CLEMSON UNIVERSITY                            *
*                               CLEMSON S.C.                               *
*                                                                          *
****************************************************************************

ENTER NUMBER OF CHARACTER TYPES.
?3


ENTER  3, 8 - CHARACTER ID'S FOR CHARACTER TYPES, ONE PER LINE.
?good
?fair
?poor
ENTER FILENAME?histdata



ENTER "LIST" - TO LIST DATA
      "SIML" - TO RUN SIMULATION
      "PRIC" - TO PERFORM PRICING
      "STOP" - TO END PROCESSING
?siml

ENTER RANGE FOR NUMBER OF BASKETS.
?2 4

ENTER SUBPOPULATION SIZE.
?48

THERE ARE THREE SUBPOPULATION SELECTION OPTIONS:
         1. SIMPLE RANDOM SAMPLING
         2. STRATIFIED RANDOM SAMPLING - PROPORTIONAL ALLOCATION
         3. STRATIFIED RANDOM SAMPLING - USER ALLOCATION
WHICH OPTION DO YOU PREFER? (1,2 OR 3)
?1

ENTER NUMBER OF REPLICATIONS.
?500


RUN STATISTICS

    POPULATION SIZE =  85

         good      =  22
         fair      =  45
         poor      =  18


    SUBPOPULATION SIZE =  48


    SELECTION METHOD:  SIMPLE RANDOM SAMPLING.

    NUMBER OF SELECTIONS CAPABLE OF CAUSING IMBALANCE =    0

    NUMBER OF REPLICATIONS =     500
```

Referring to this printout, note that access to the simulation program was obtained by issuing the command "run basket*" and responding with a "1" to the first question.

Following the printing of a header, the user is prompted for the number of character types. If the baskets are to be balanced on bid price only, all proposals are of the same "type" so the response should be a "1". Otherwise, enter the integer from 2 to 12 which corresponds to the number of different proposal types. In the printout, we enter 3 corresponding to the use of 3 characteristics.

The user is then prompted to supply names (up to 8 characters per name) for the various "types" of proposals. The first name supplied will be associated with proposals identified on the input file as type 1, the second name supplied will be associated with type 2 proposals, and so on. Of course, if all proposals are the same "type", this question will not be asked. Here we supply 3 names: good, fair, and poor.

Next, the user is prompted for the name of the input file containing the historical data which, in this example, is "histdata".

Four options are then provided:

(1) LIST: This will cause the computer to list the contents of the input file in convenient tabular form showing bid price, negotiated price and the name of the proposal type. It is recommended that this option be exercised prior to running the simulation program as a check on the accuracy of file input. It was not done here in order to save space.

(2) SIML: This causes the computer to run the perspective analysis simulation program.

(3) PRIC: This causes the computer to make a <u>single</u> pass through the simulation program resulting in a printout of basket formation, estimated prices, and error due to sampling.

(4) STOP: This returns control to the "READY" mode. Referring to the printout, note that the "siml" option was selected.

The user is then prompted for a "RANGE FOR THE NUMBER OF BASKETS" desired. The response should be 2 integers separated by a space. For example, the response 2 6 will result in error histograms being produced for 2,3,4,5 and 6 basket setups. If one were only interested in looking at the error distribution associated with 25% sampling (4 baskets), the response should be 4 4. The second integer must be at least as large as the first but no larger than the current program limit of 10.

Next, the user is prompted for a subpopulation size. Guidelines for providing the response to this question are discussed in Part C.

The prompt dealing with "SUBPOPULATION SELECTION OPTIONS" refers to the method used to select populations of the specified size from the historical data set (superpopulation). Three alternatives are given:

(1) SIMPLE RANDOM SAMPLING: Populations of the specified size are selected at random from the superpopulation without regard to character type. This method is appropriate for most pricing applications and was used in our example.

(2) STRATIFIED RANDOM SAMPLING: PROPORTIONAL ALLOCATION: Populations of the specified size are selected in such a way that each selected population will have the same proportion of each type of proposal as exists in the superpopulation of historical data. This option is appropriate when it is known in advance that the proposal "stream" which will be subjected to sampling

will bear a <u>very</u> close resemblance to the superpopulation in terms of the breakdown by proposal type.

(3) STRATIFIED RANDOM SAMPLING: USER ALLOCATION: This population selection method is similar to option (2) except that the proportions are specified by the user instead of conforming automatically to those of the superpopulation. This option is appropriate when the Basket Method is to be used on a "one-shot" basis in estimating the price of a single well-defined backlog of proposals. In that case, every generated population can be forced to look exactly like the target backlog in terms of how <u>many</u> proposals of each type will be selected.

Finally, the user is prompted for the number of replications. This refers to the number of different populations that will be selected from the superpopulation of historical data and subjected to basket sampling as discussed in Part C. Experience has shown that 300 to 500 replications is sufficient for most situations. "Bigger is better" in terms of accuracy but also more expensive in terms of computer time. Computing during "non-prime-time" or, better yet, use of the IND overnight batch option is recommended. This option is discussed in the G. E. Command System Reference Manual.

The error frequency histograms resulting from the above example of the simulation program appear on page 33.

<u>Example (2): "Basket Selection Model</u>".

Prior to running the basket selection routine, the user must create a file of all population bid prices as input. The format for a single line of input for this type of data set consists of a bid price (expressed in $1000's) and, if used, a characteristic type (an integer). A comma is used to separate the two numbers. In the example input below, dots (•••) have been inserted for brevity instead of showing the entire input. Assume the user has selected the filename "bidfile" for this data set. After the logon process is completed and the computer terminal is in the "READY" mode, commands are issued and data entered as follows (upper case lines are prompts by the computer). The DEM mode is terminated by depressing the "break" key.

```
READY
new bidfile

READY
dem

READY FOR INPUT
922,2
887,3
884,3
876,2
854,1
  .
  .
  .
128,2
127,1
122,1
119,1

READY
save

READY
```

Having created this file of bid prices, balanced baskets are created as discussed in Part B by running the "Basket Selection Model". In the course of running this program, the user is prompted to provide necessary input by a question mark (?) following a brief description of what input is needed.

```
READY
run basket*

BASKET*     12:35PST    12/11/81

WHICH OF THE FOLLOWING WOULD YOU LIKE TO RUN:

  1) BASKET SIMULATION MODEL
  2) BASKET SELECTION MODEL
  3) AWARD MODEL
  4) TERMINATE

INPUT 1,2, OR 3
 ?2
**************************************************************************
*                                                                        *
*                            BASKET METHOD                               *
*                      VERSION III - JUNE 1981                           *
*                                                                        *
*            THEORETICAL DEVELOPMENT - DR. K. T. WALLENIUS               *
*        ALGORITHM DESIGN AND IMPLEMENTATION - DR. STEPHEN BENZ          *
*                                                                        *
*                          CLEMSON UNIVERSITY                            *
*                             CLEMSON S.C.                               *
*                                                                        *
**************************************************************************
ENTER NUMBER OF BASKETS.
?4

ENTER NUMBER OF CHARACTER TYPES.
?3

ENTER  3, 8 - CHARACTER ID'S FOR CHARACTER TYPES, ONE PER LINE.
?good
?fair
?poor

ENTER FILENAME?bidfile

RUN STATISTICS

     NUMBER OF BASKETS =             4
     NUMBER OF CHARACTER TYPES =     3
     NUMBER OF BIDS =               48
     NUMBER OF BIDS PER CHARACTER TYPE:

              1.  good      =   22
              2.  fair      =   16
              3.  poor      =   10
 ---------------------------------------

BASKET NO.   1
        PROPOSAL NO.             BID PRICE          CHARACTERISTIC

       1.      6                  854.000              good
       2.     13                  750.000              good
       3.     23                  479.000              good
       4.     30                  281.000              good
       5.     38                  191.000              good
       6.     12                  782.000              fair
       7.     25                  361.000              fair
       8.     29                  284.000              fair
       9.     45                  128.000              fair
      10.      2                  887.000              poor
      11.     27                  296.000              poor
      12.     33                  224.000              poor

BASKET NO.   2
        PROPOSAL NO.             BID PRICE          CHARACTERISTIC

       1.      9                  817.000              good
       2.     18                  573.000              good
       3.     19                  561.000              good
       4.     28                  291.000              good
       5.     44                  129.000              good
       6.     48                  119.000              good
       7.      1                  922.000              fair
       8.     17                  651.000              fair
       9.     36                  203.000              fair
      10.     41                  144.000              fair
      11.      3                  885.000              poor
      12.     34                  223.000              poor

BASKET NO.   3
        PROPOSAL NO.             BID PRICE          CHARACTERISTIC

       1.     11                  786.000              good
       2.     15                  660.000              good
       3.     21                  514.000              good
       4.     26                  297.000              good
       5.     43                  139.000              good
       6.      5                  876.000              fair
       7.     20                  541.000              fair
       8.     35                  215.000              fair
       9.     40                  150.000              fair
      10.      4                  884.000              poor
      11.     31                  269.000              poor
      12.     39                  186.000              poor

BASKET NO.   4
        PROPOSAL NO.             BID PRICE          CHARACTERISTIC

       1      10                  795.000              good
       2.     14                  685.000              good
       3.     22                  506.000              good
       4.     32                  237.000              good
       5.     46                  127.000              good
       6.     47                  122.000              good
       7.      7                  847.000              fair
       8.     24                  378.000              fair
       9.     37                  196.000              fair
      10.     42                  143.000              fair
      11.      8                  818.000              poor
      12.     16                  664.000              poor
 ----------------------------------------

BASKET STATISTICS

     DIFFERENCE BETWEEN MAX AND MIN BASKETS =        1.000
     MEAN ERROR BOUNDED BY-0.01% AND 0.01%.
     BID PRICE MOMENTS:

       BASKET    NO. PROPOSALS    TOTAL OF BIDS      MEAN      STD DEV
         1            12            5517.000       459.750    280.212
         2            12            5518.000       459.833    310.051
         3            12            5517.000       459.750    286.971
         4            12            5518.000       459.833    291.748
       POPUL          48           22070.000       459.792    282.967
```

After issuing the command "run basket*" and requesting the "Basket Selection Model" by entering a "2" when requested to pick one of the 3 programs to run, the user is prompted for the number of baskets desired. Suppose, based on simulation results and other relevant considerations, it has been decided that 4 baskets will be used (i.e., a 25 percent sample). This information is given the computer in response to the prompt "ENTER NUMBER OF BASKETS". In response to the prompt "ENTER NUMBER OF CHARACTER TYPES", the user enters the number of different character types, 3 in this example.

The user is then prompted to supply names, up to 8 characters per name, for the various types of proposals as explained in example (1). The answer to the prompt "ENTER FILENAME" tells the computer where to look for the file of bid prices which it will then partition into balanced baskets. That name is "bidfile" in this example.

Examining the output, we note that the contents of "bidfile" are first summarized under the heading "RUN STATISTICS". This is followed by a printout of the contents of each basket by proposal number (corresponding to the order of the data in the input file), the bid price, and the type of proposal. Basket contents are summarized under the heading "BASKET STATISTICS".

At this point in an actual application, one of the baskets would be selected at random and its contents subjected to the prescribed analysis and negotiation process. The formulas given in Part B can easily be applied to estimate negotiated prices for proposals in the remaining baskets or the "Award Model" program can be used to perform these computations and provide a convenient tabular output. This program will now be discussed.

Example (3): "AWARD MODEL".

Prior to running this model, the user must create a file of proposal numbers and negotiated prices for use as input. The first line of this file must be the name of the file containing the bid prices used as input to the "BASKET SELECTION MODEL" program. The format for each subsequent line is a proposal number (the sequence number assigned the proposal by the computer when the baskets were formed) and the negotiated price for that proposal (expressed in $1000's). Assume the user selects the name "negfile" and that "bidfile" is the name of the file used as input to the program which created the baskets. After the logon process is completed and the computer terminal is in the "READY" mode, commands are issued and data entered as before (upper case lines are prompts by the computer). The DEM mode is terminated by depressing the "break" key. In this example, we assume basket 3 of the previous output was selected at random and its proposals analyzed and negotiated in the prescribed manner. Proposal numbers and negotiated prices are entered as follows.

```
READY
new negfile

READY
dem

READY FOR INPUT
bidfile
11,752
15,629
21,491
26,282
43,133
5,792
20,486
35,195
40,134
4,679
31,182
39,126

READY
save

READY
```

Having created this file of proposal numbers and negotiated prices, price computations for the unsampled proposals are obtained by running the "AWARD MODEL" as shown below. This program simply computes the decrement ratio R and applies it to unsampled proposals to obtain estimated prices. Proposals which were actually negotiated can be identified by an entry in the column headed "NEGOTIATED PRICE". The rest of the output is self explanatory.

```
READY
run basket*

BASKET*     09:15PST    12/04/81

WHICH OF THE FOLLOWING WOULD YOU LIKE TO RUN:

  1) BASKET SIMULATION MODEL
  2) BASKET SELECTION MODEL
  3) AWARD MODEL
  4) TERMINATE

INPUT 1,2, OR 3
 ?3


********************************************************************************
*                                                                              *
*                            PRICE COMPUTATIONS                                *
*                        BASED ON NEGOTIATED SAMPLE                            *
*                                                                              *
********************************************************************************

ENTER NAME OF FILE CONTAINING PROPOSAL NUMBERS AND NEGOTIATED PRICES ?negfile
```

PROPOSAL NUMBER	BID PRICE	ESTIMATED PRICE	NEGOTIATED PRICE
1	922.000	815.712	
2	887.000	784.747	
3	885.000	782.977	
4	884.000		679.000
5	876.000		792.000
6	854.000	755.551	
7	847.000	749.358	
8	818.000	723.701	
9	817.000	722.816	
10	795.000	703.352	
11	786.000		752.000
12	782.000	691.851	
13	750.000	663.540	
14	685.000	606.033	
15	660.000		629.000
16	664.000	587.454	
17	651.000	575.953	
18	573.000	506.945	
19	561.000	496.328	
20	541.000		486.000
21	514.000		491.000
22	506.000	447.668	
23	479.000	423.781	
24	378.000	334.424	
25	361.000	319.384	
26	297.000		282.000
27	296.000	261.877	
28	291.000	257.454	
29	284.000	251.260	
30	281.000	248.606	
31	269.000		182.000
32	237.000	209.679	
33	224.000	198.177	
34	223.000	197.293	
35	215.000		195.000
36	203.000	179.598	
37	196.000	173.405	
38	191.000	168.982	
39	186.000		126.000
40	150.000		134.000
41	144.000	127.400	
42	143.000	126.515	
43	139.000		133.000
44	129.000	114.129	
45	128.000	113.244	
46	127.000	112.359	
47	122.000	107.936	
48	119.000	105.282	
	---------	---------	---------
TOTALS	22070.000	14644.770	4881.000

REFERENCES

(1) Royall, Richard M. and Herson, Jay. (1973) Robust Estimation in Finite Populations I. Journal of the American Statistical Association, 68, 880-9.

(2) Wallenius, K.T. (1980) Statistical Methods in Sole Source Contract Negotiation. Journal of Undergraduate Mathematics and its Applications, Vol. 1, No. 2, 113-23.

(3) Wallenius, K.T. (1981) The Basket Method for Selecting Balanced Samples, Part I: Theory. Technical Report N129, Department of Mathematical Sciences, Clemson Univeristy.

(4) Wallenius, K.T. and Benz, Stephen. (1981) The Basket Method for Selecting Balanced Samples, Part III: Computer Codes. Technical Report N134, Department of Mathematical Sciences, Clemson University.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N 132	2. GOVT ACCESSION NO. AD A112 949	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Basket Method for Selecting Balanced Samples Part II: Applications to Price Estimation		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER TR #377
7. AUTHOR(s) K. T. Wallenius		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0451
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clemson University Dept. of Mathematical Sciences Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 365-049
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Code 434 Arlington, Va. 22217		12. REPORT DATE December 1981
		13. NUMBER OF PAGES 53
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Balanced subsampling; Price estimation; Perspective analysis.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The "Basket Method" of sampling, a tool designed to achieve statistically balanced samples, is described in intuitive terms. Special reference is made to applications in price analysis where experience has demonstrated the practicality of the technique. The intent is to provide an overview of what the system is intended to do and how it does it in order to assist price analysts and negotiators expedite proposal processing while maintaining acceptable levels of risk. Guidelines and examples are given

(continued on back

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) for implementing a statistical pricing program tailored to local conditions. Underlying theory and documented computer codes are provided separately in Part I and Part III, respectively.

EN

DAT
FILM

5-

DT